DOCUMENT RESUME

ED 308 613 EA 021 113

AUTHOR Mandeville, Garrett K.
TITLE An Evaluation of PET Based on Longitudinal Data.
PUB DATE Mar 89
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Elementary Secondary Education; *Inservice Education; *Longitudinal Studies; *Models; *Program Evaluation; Staff Development; *Teacher Effectiveness
IDENTIFIERS Coaching; *Program for Effective Teaching; *South Carolina

ABSTRACT

Although teacher inservice programs based on Madeline Hunter's Program for Effective Teaching (PET) have become very popular in U.S. schools, there is little evidence that the Hunter model ultimately results in increased student achievement. This longitudinal study attempts to evaluate the effects of Hunter-based staff development programs on student achievement. The study is based on Mandeville's previous research involving 900 PET-trained South Carolina teachers, only 26 of whom had completed training 2 years prior to the school year for which student achievement data were gathered. (Hunter has claimed that to reach the "artistic" level, PET-trained teachers minimally require a two-year growth process of study, articulation, practice, and internalization, supported by coaching, classroom observation, and followup conferences.) Since 26 is a relatively small sample size, the current study began with 123 second- and third-grade teachers who were PET-trained during the second implementation year (1984-85). Data were also gathered for 888 second- and 727 third-grade teachers who were not PET-trained (NOPET). Test data were developed from the statewide Basic Skills Assessment Program in reading and math. The PET and NOPET teachers were compared using a number of variables (gender, ethnicity, age, and years of teaching experience). Classroom level data for 4 years (1984-85 to 1987-88) were also compared for the two groups; The final PET sample was 44; the final NOPET sample was 34. Since the PET and NOPET teacher samples were comparable, the reasonably random nature of each group's average profile indicates that no overall PET effect has been manifest. The increasing profiles anticipated as a teacher moved toward artistry did not occur, except incidentally. Perhaps coaching quality is to blame for the results. Sixteen references and several data tables are included. (MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308613

An Evaluation of PET Based on Longitudinal Data

Garrett K. Mandeville

University of South Carolina

Running Head: A PET EVALUATION USING LONGITUDINAL DATA

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Garrett K. Mandeville

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, March 1989

EA 021 113

Teacher inservice programs based on the work of Madeline Hunter have become very popular in the U.S. over the past decade. Unfortunately, as has been noted by Slavin (1987) and Gibboney (1987), little evidence has been provided that implementation of the Hunter Model ultimately results in increased student achievement. This study attempts to add to our knowledge of the effects of Hunter-based staff development programs on student achievement using a longitudinal design. It is an extension of work by the writer [Mandeville (1988a) and Mandeville & Rivers (1988)] in which no positive (or negative) effects have been detected.

Background

Among the studies of this topic which did use student test data was the Napa/Vacaville Project (Stallings & Krasavage, 1986) but limitations in the evaluation design rendered the results inconclusive. Dildy's (1982) study did produce positive effects but was of such a limited nature that the findings should be interpreted with caution. The SIM Project (Manatt & Stow, 1986) included two teacher inservice options, one of which (Essential Elements of Instruction) is based on Hunter's work. Unfortunately, an adequate control group existed in only one of the six study sites (Minneapolis) and less than one-third of the teachers opting for inservice training chose EEI. Thus, the fact that no significant positive effects of the inservice experience on student achievement were detected provides little evidence about EEI. Donovan, Sousa and Walberg (1987), reporting on a Hunter-based staff developement program in West Orange New Jersey, also found no difference in achievement or grades for students of participating and nonparticipating teachers.

The context of Mandeville's (1988a) study was the South Carolina implementation of the Program for Effective Teaching (PET) which was begun in the Spring of 1984. A unique focus of this study was that the design took into consideration Hunter's belief that the effectiveness of most teachers will not be materially changed as a result of the relatively brief initial training which is involved in most Hunter-based staff development programs. Rather, she asserts that "For most educators, progressing from Phase I to Phase IV is a minimum two year growth process of study, articulation, practice, and internalization." (Hunter and Russell, undated). Essentially, when a teacher reaches the "artistic" Phase IV teaching level he or she makes decisions which are contingent (or conditioned) by the teacher-student interactions which have preceded. Crucial to this growth process is coaching which involves formative classroom observations and follow-up conferences, usually conducted by the principal (who has also been PET trained).

There is little empirical evidence to support or refute Hunter's opinion on this matter. At least one study produced anecdotal evidence which was somewhat supportive, however: "Essential Elements took several years to master according to staff members at Northfield who had much experience with the process." (Manatt & Stow, 1986, p. 121).

Although 900 PET trained teachers were included in the Mandeville (1988a) sample, only 26 satisfied the requirement of having completed training two years prior to the school year for which the student achievement data were gathered (1986-87). Since 26 is a relatively small sample size and there is always the possibility that the first year cohort of trainees may have

been atypical, the current study began with 123 of the teachers who were PET trained during the second year of implementation, i.e., between September 1984 and August 1985. Thus, these teachers had an opportunity to be coached during the 1985-86 and 1986-87 school years so the scores of their students at the end of the 1987-88 school year could be expected to reflect a year of artistic teaching. In addition, since improvement in achievement might conceivably occur at any point over the four year span, it seemed worthwhile to gather data to allow a characterization of how these changes occurred. Thus, the performance of each sample teachers' students for each of these four school years was obtained.

Methods

This study began with a sample of 69 2nd grade and 54 3rd grade teachers who completed PET training prior to the beginning of the 1985-86 school year. On the same data set were 888 2nd grade and 727 3rd grade teachers who were not PET trained (NOPET) prior to the drawing of the sample for Mandeville (1988a).

To reduce the labor involved in developing the longitudinal data sets, random samples (by grade) of the NOPET teachers were drawn which were roughly twice as large as the PET sample sizes (n=140, 2nd grade and n=110, 3rd grade). The initial NOPET sample sizes were purposefully made larger than the corresponding sample sizes for PET teachers because it was anticipated that many of these teachers had subsequently received PET training. The total sample size at this point was 373.

Since the Mandeville (1988a) study involved only 1987 posttest data, the next task was to locate the test performance of the students of each teacher in the sample for the previous

two school years (1984-85 and 1985-86) and the following year (1987-88). To facilitate the discussion, it will be useful to briefly describe two relevant South Carolina educational initiatives: a testing program and a school reward program.

Testing Program

South Carolina currently tests each public school student in the 1-11 grade span with either a locally developed criterion referenced test under the Basic Skills Assessment Program (BSAP) or a commercially produced norm referenced test under the Statewide Testing Program. Germaine to this study is the fact that at the 1st, 2nd and 3rd grades, BSAP tests assessing reading and mathematics skills are used. It is important to note that the BSAP tests were developed to test state objectives in reading and math so that they should rank high on curricular validity. (Department of Education audit teams have verified this in about 90% of South Carolina's schools.)

School Incentive Reward Program (SIRP)

One of the various ways these statewide test data are utilized is in the allocation of approximately $4 to $7 million annually to schools under the SIRP. The primary ingredient in these allocations is a School Gain Index (SGI) which is based on the performance of all students in each school in the 1-11 grade span who were tested. Each student's performance is operationalized by the (standardized) discrepancy between his or her actual scores in reading and mathematics and those predicted from a statewide regression analysis. In preparation for the regression analyses, the new test records are matched on an individual (student) basis with those from the previous year (e.g., Spring 1988 2nd grade test records are matched with Spring

1987 1st grade records). Subsequent to regression analyses, files containing each student's residual scores are created. These data files were created for the 1986, 1987, and 1988 SIRP analyses using the Spring test data for each of those years. In addition, comparable data sets for the 1984-85 school year were created during a feasibility study of the proposed SIRP analytic approach.

Matching Student Data to Teachers

Among the various data elements retained on the matched student data sets produced during the SIRP analyses were the "NAMEs" from the Classroom Header record which precedes each set of student records. In most cases, the NAME is the name of the teacher who monitored the administration of the test and--if the class is self-contained--it usually refers to the child's teacher. Thus, when student residuals are aggregated to the NAME level, they provide an index of how well that teacher's students did relative to the prediction. In this paper, the average residuals will be referred to as Teacher Gain Indices (TGIs) and one will exist for reading (TGIR) and math (TGIM). The TGIs based on the Spring 1987 testing were used in the Mandeville (1988a) paper and, as noted above, this data set served as the starting point for the development of the longitudinal data base. For more detail on the safeguards imposed to assure that the TGIs used in Mandeville (1988a) correctly reflected the performance of the students of a given teacher and that the teacher was responsible for reading and math instruction, see the earlier paper.

For the other three years, the student level files were also aggregated to the NAME level. Next, each record was matched by

NAME against the teacher names on the source file containing the Spring 1987 data. The program used during this matching process was a modification of one written by the author and described in Mandeville (1988b). During this record linking, matching records were not found for all of the 373 source records. The reasons for this are manifold. Some are: 1) teacher left the school district via retirement, death, termination, etc.; 2) teacher was reassigned to a different grade level or a nonteaching position; or 3) the record linking program did not identify the match, possibly due to a name change. The smallest number of valid matches was for the Spring 1985 test data where 236 of the 373 (63%) were identified. For the other two years, over 70 percent of the 373 records were correctly located. When the four files were merged, the number of complete records (data for all four years) was 163, representing 69 PET and 94 NOPET teachers.

The Teacher Survey

At this point the data set was clean with two possible exceptions. First, since the self-contained status of the classes was based on the 1986-87 school year, it was possible that it was not appropriate for one or more of the other three years. If so, this would render the data attribution scheme for the year(s) in question to be invalid. Second, because of the popularity of PET, it was anticipated that a number of the NOPET teachers had received training after February 1987 when the information on PET trainees was provided to the researcher.

To address these two issues, forms were sent to the 163 teachers requesting that they validate or correct this information. For PET teachers, the form included a listing of the schools (as identified based on the test data) "that they

taught in" during the previous four years and a question to verify that they had completed first cycle PET training prior to the beginning of the 1985-86 school year. For any of the four years the teachers were to indicate if the class was not self-contained. For NOPET teachers, the information concerning their teaching schedules for the past four years was the same, but they were to indicate whether they had been PET trained and, if so, when first cycle training had been completed.

Data Analyses

Initially, the PET and NOPET teachers were compared on a number of background variables of interest. The teacher level variables included were gender, ethnicity, age, and years of teaching experience. The following classroom level data for each of the four years were also compared for the two teacher groups: 1) the percentage of children eligible for free or reduced price lunches; 2) the percentage of children repeating the grade; 3) class size; and 4) previous student test performance (measured the prior spring). To maintain reasonable sample sizes (and therefore statistical power), the analyses for the variables described in 1-3 above were conducted for the combined data set, i.e., both 2nd and 3rd grade teachers. This was not be possible with previous test performance since the BSAP2 and BSAP3 are not in a common metric.

The analyses to address the primary research question involved the use of the four TGIs in each of the two subject areas. For example, regarding reading TGIR85, TGIR86, TGIR87, and TGIR88 were available for each teacher. The analytic approach was to consider these data in a repeated measures design (analogue of the split-plot) with "year" as the within subjects

(teachers) factor and treatment (PET vs. NOPET) as the between subjects factor. If the achievement of the students taught by the PET teachers showed an unusual increase at some point in time, the hypothesis of no treatment by year interaction should be rejected. Because the standardized residuals statewide can be considered to be in a common metric across grades (approximately unit normal), this analysis was conducted on the combined data set.

Results

Survey Returns

One month after the mailing, 46 of the 69 PET teachers (67%) and 51 of the 94 NOPET teachers (54%) had returned the forms. Although follow-up telephone calls were contemplated, preliminary data analyses suggested that this would be unnecessary. All PET teachers indicated that they had been trained but one reported being on leave for the last part of the 1987-88 school year and another indicated that her class had not been self-contained during one of the years so the PET teacher sample size was reduced to 44. As expected, more than one-half of the responding "NOPET" teachers had participated in PET training. However, the majority were trained after the Spring 1988 testing so they could be considered NOPET for purposes of the study. Only 15 were lost for the reason of PET training prior to Spring, 1988, along with two who indicated problems with the assumption that there classes were self-contained. Thus, the NOPET sample size was 34.

Comparisons on Background Variables

The PET and NOPET teachers were found to be very similar (and not statistically different) on gender (100% female), ethnicity (88.5% white), age (M = 39.2 years), and years of

teaching experience (M = 12.5). Nor were there differences for any of the four years on the percentage of children eligible for free or reduced price lunches which ranged from about 38% to 42% on the average; for the percentage of children repeating the grade which averaged between 7% and 10% for the 1984-85 school year and 3% to 5% thereafter; or for class size which was between 21 and 22 students on the average.

Previous student test performance was compared for reading and math for each of the four years, thus eight comparisons for each grade level. Of these 16 t-tests, one was borderline significant ($t = 2.02$, d.f. = 40, $p = .0503$) favoring the students of 2nd grade NOPET teachers in math for the 1984-85 school year (actually, performance on the BSAP1 in the Spring of 1984). The mean difference was about 20 scale score points. Since the hypothesis of interest focusses on the performance of students of PET teachers in more recent years (especially 1987-88) this one borderline significant difference was not considered to be a serious flaw in the research. None of the other differences approached statistical significance and, at the descriptive level, some favored each of the two teacher groups.

ANOVA Results for Main Questions

The repeated measures ANOVA results for reading are presented in Table 1. Whether the multivariate or univariate

---

Insert Table 1 about here

---

approach to the test for an interaction of year and treatment was used made no difference--all F ratios are less than 1.0 in value. In fact, sphericity appears to hold quite well using Mauchly's

criteria (.934 with a p-value for the associated chi square of .40)--so the univariate Fs are probably dependable. The tests of the repeated measures factor, years, although not relevant to the discussion, are included for completeness.

In this model, the correlations among the repeated measures are expected to be sufficiently large so that the error term appropriate for testing within unit (here teacher) effects is small. Here, however, the correlations are not large, ranging between .3 and .4 with an average r of .34. The ANOVA approach to estimating the reliabiliity of the four-year average yielded .65 (see Winer, 1971, p. 287); applying the Spearman-Brown formula to the average r of .34 produced a comparable .67. Thus, the reading performance of the students of these teachers was not, in general, very consistent over the four years.

Parallel results for mathematics are provided in Table 2.

Insert Table 2 about here

Here the F ratios tend to be larger but none achieves significance at the common .05 level. For mathematics more stability is evident since the average of the six cross-year rs was .50 which translates to a reliability coefficient of .80 for the four-year composite. The cyclical pattern of the mean differences (the N-P row of Table 2), although probably meaningless, did cause the cubic component of the interaction to be statistically significant ($F = 4.16$, $p = .045$). Therefore, when aggregated across the two groups, the average profiles in reading and the average profiles in math do not exhibit variation beyond than expected by chance.

Individual Teacher Profiles

Given that no overall effects were manifest, the writer conducted analyses to identify unusual individual teacher profiles. This analysis was done prior to the inclusion of about the last ten teachers in the sample but, since the profiles are presented purely for descriptive purposes, this should not be considered a drawback.

Thus, for the 68 cases involved (38 PET and 30 NOPET), the four teachers with the largest average increase in reading over the four year period were identified. The other subject area (math) was treated similarly. The next set of profiles considered were those with the largest (in absolute value) decreases from the 1984-85 school year to 1987-88. Again both reading and math were treated separately.

The four teachers with the largest increases in reading are shown in Figure 1. Interestingly, all four teachers were PET

Insert Figures 1-4 about here

---

trained. Two (PET#224 and PET#348) exhibit rather strong increasing trends over the last two years. This would be a reasonable type of thing to expect as the teacher became more able to draw on her repetoire of skills more automatically, i.e., achieved the artistic teaching level. One could also contemplate that PET#19 achieved that level more rapidly and then tended to level-off. However, the profile for the teacher with the largest increase, PET#306, does not support this type of thinking.

When we turn to the four largest increases in mathematics as

displayed in Figure 2 we see one "repeater", PET#19, but her math profile differs considerably from her reading profile, showing a decrease for the last year. In fact, the two NOPET entries show more continuous progress than is exhibited by either of the PET entries.

On the other side of the coin, consider the four largest decreases in Figure 3. Here, three of the four teachers were PET trained and, by-and-large, the decreases appear rather consistent over the four years. Not what one would hope for, irrespective of the type of staff development support a teacher was getting. Figure 4 with math decreases is not so regularly patterned and contains three NOPET teachers.

Discussion

Since the PET and NOPET teacher samples were quite comparable, the reasonably random nature of the average profiles for the two groups indicates that no overall PET effect has been manifest. The increasing profiles which were anticipated as a teacher moved towards artistry simply did not manifest themselves, except possibly incidentally. For a believer in PET, the blame will likely be placed on the amount and/or quality of the coaching which is supposed to occur during the (up to) two year practice period. In fact, the results presented in Mandeville and Rivers (in press) support this contention: only 57% of the PET teachers in that study received a very minimal amount of coaching and the coaching received by one-third to one-half of the coaching recipients deviated from Hunter recommendations in important ways. It was unfortunate that data on coaching was not gathered in this study for it might have shed some light on the individual teacher profiles presented.

That the "teacher performance" measures used in this paper were more stable in math than reading is becoming more-or-less fact. The same phenomena occurred in the context of the stability of similar types of indices at the school level (Mandeville & Anderson, 1987; Mandeville, 1988c). Also, in Mandeville and Rivers (in press) modest effects associated with PET were manifest for math but not for reading. The suggested mechanism is that math is more related to "school learning" than reading.

It was reported that roughly one-half of the teachers in South Carolina were trained prior to April, 1988 (Tudor, 1988). Based on the responses of our "NOPET" sample, it is possible to estimate that about 21% more of the state's teachers received 1st cycle training between April, 1988 and February, 1989. Thus, it appears as the emphasis on training teachers has not diminished.

It is generally agreed upon that coaching is at least one of the primary problems with the state PET implementation, possibly the most important problem (see Hunter, 1988). What is not clear is how an increase in the number of teachers who require ample amounts of quality coaching is going to interact with the coaching problem. Unless more and/or improved coaches with more time to conduct observations are being produced through additional training or inservice supervision, the problem could be exacerbated.

References

Dildy, P. (1982). Improving student achievement by appropriate teacher in-service training: Utilized Program for Effective Teaching (PET). Education, 103(2), 132-138.

Donovan, J. F., Sousa, D. A., and Walberg, H. J. (1987). The impact of staff development on implementation and student achievement. Journal of Educational Research, 80(6), 348-351.

Gibboney, R. A. (1987). A critique of Madeline Hunter's teaching model from Dewey's perspective. Educational Leadership 44(5), 46-50.

Hunter, M., and Russell (undated). Critical aspects of a staff development program to increase instructional effectiveness.

Hunter, M. (1988). "Well acquainted" is not enough: A response to Mandeville and Rivers. Educational Leadership, 46(4), 66-67.

Manatt, R. P. and Stow, S. B. (1986). Developing and testing a model for measuring and improving educational outcomes of K-12 schools. (Technical Report). Ames, Iowa: Iowa State University, School Improvement Model (SIM), College of Education.

Mandeville, G. K. (1988a, April). An evaluation of PET using extant achievement test data. In R. Ginsberg, (Chair), South Carolina's implementation of the Hunter model: Issues and answers. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans.

Mandeville, G. K. (1988b, April). A record linking system for research on school improvement. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Mandeville, G. K. (1988c). School effectiveness indices revisited: Cross-year stability. Journal of Educational Measurement, 25(4), 349-356.

Mandeville, G. K. and Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. Journal of Educational Measurement, 24(3), 203-216.

Mandeville, G. K. and Rivers, J. (1988). Effects of South Carolina's PET program. Educational Leadership, 46(4), 63-66.

Mandeville, G. K. and Rivers, J. (In press). Is the Hunter model

a recipe for supervision? Educational Leadership.

Slavin, R. (1987). The Hunterization of America's schools. Instructor, 96(8), 56-60.

Stallings, J. and Krasavage, E. M. (1986). Program implementation and student achievement in a four-year Madeline Hunter follow-through project. The Elementary School Journal, 87(2), 117-138.

Tudor, J. D. (1988, April). The program for effective teaching in South Carolina. In R. Ginsberg, (Chair), South Carolina's implementation of the Hunter model: Issues and answers. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans.

Winer, B. J. (1971) Statistical Principles in Experimental Design. New York:McGraw-Hill

Table 1

ANOVA Results for Reading

	Means by Year					
	1985	1986	1987	1988	MV F Tests	UV F Tests
NOPET	1.05	1.04	-.14	.20	F_V=.67	F_V=.79
PET	.61	.26	.49	.00	F_{V*T}=.51	F_{V*T}=.56
N-P	.44	.78	-.63	.20		F_T=.08

Note: F_{V*T} refers to year by treatment interaction F ratios; F_V refers to the less relevant test regarding differences in the yearly averages; and F_T is the between treatments F ratio concerned with the equality of the four year averages for the two treatment groups.

Table 2

ANOVA Results for Math

	Means by Year					
	1985	1986	1987	1988	MV F Tests	UV F Tests
NOPET	.25	1.67	-.76	1.02	F_V=1.88	F_V=1.97
PET	-.17	.13	-.04	-.05	F_{V*T}=1.37	F_{V*T}=1.49
N-P	.42	1.54	-.72	1.07		F_T=.42

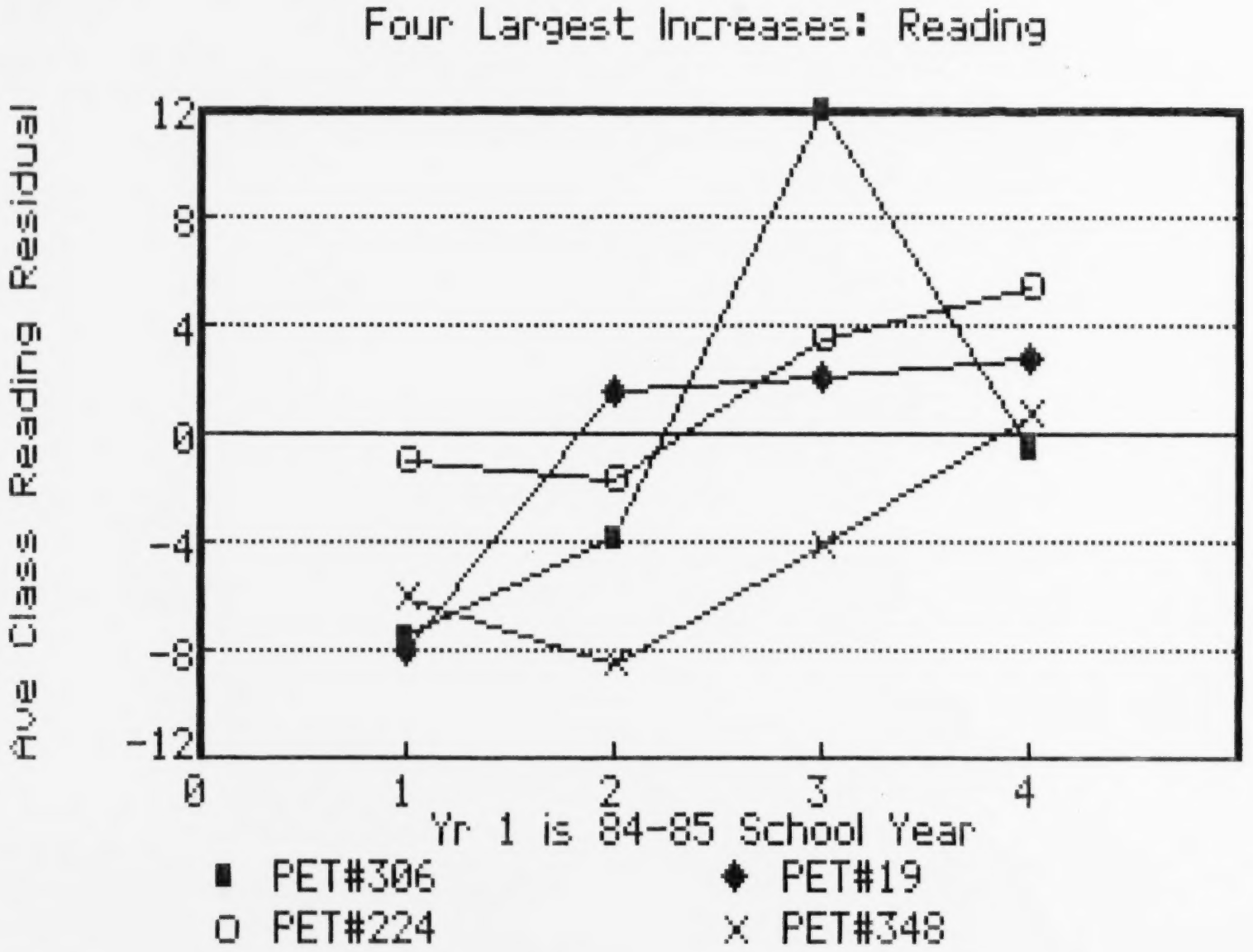
Four Largest Increases: Reading
Ave Class Reading Residual
12
8
4
0
-4
-8
-12
0
1
2
3
4
Yr 1 is 84-85 School Year
PET#306
PET#19
PET#224
PET#348

Figure 1

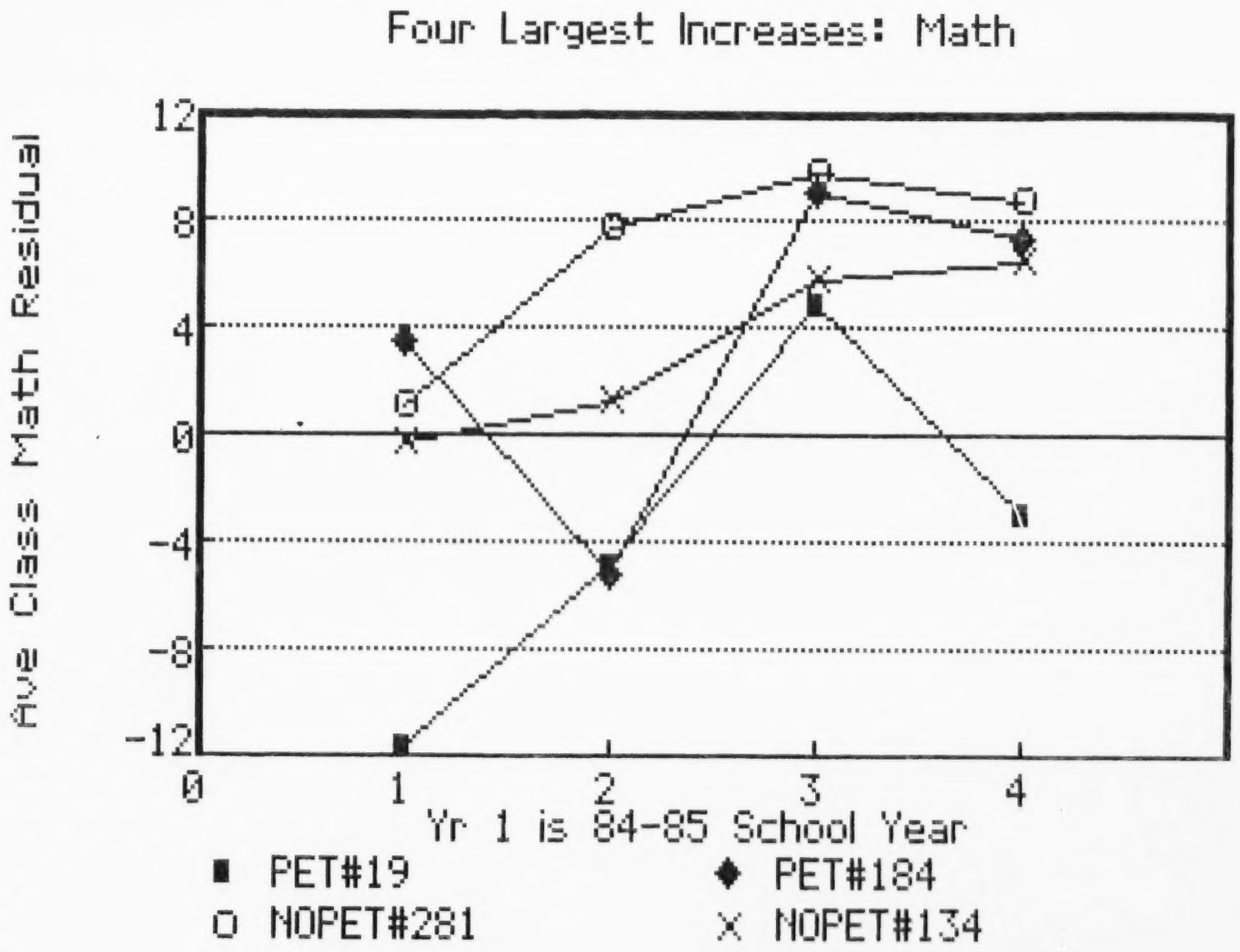
Four Largest Increases: Math
Ave Class Math Residual
12
8
4
0
-4
-8
-12
0
1
2
3
4
Yr 1 is 84-85 School Year
PET#19
PET#184
NOPET#281
NOPET#134

Figure 2

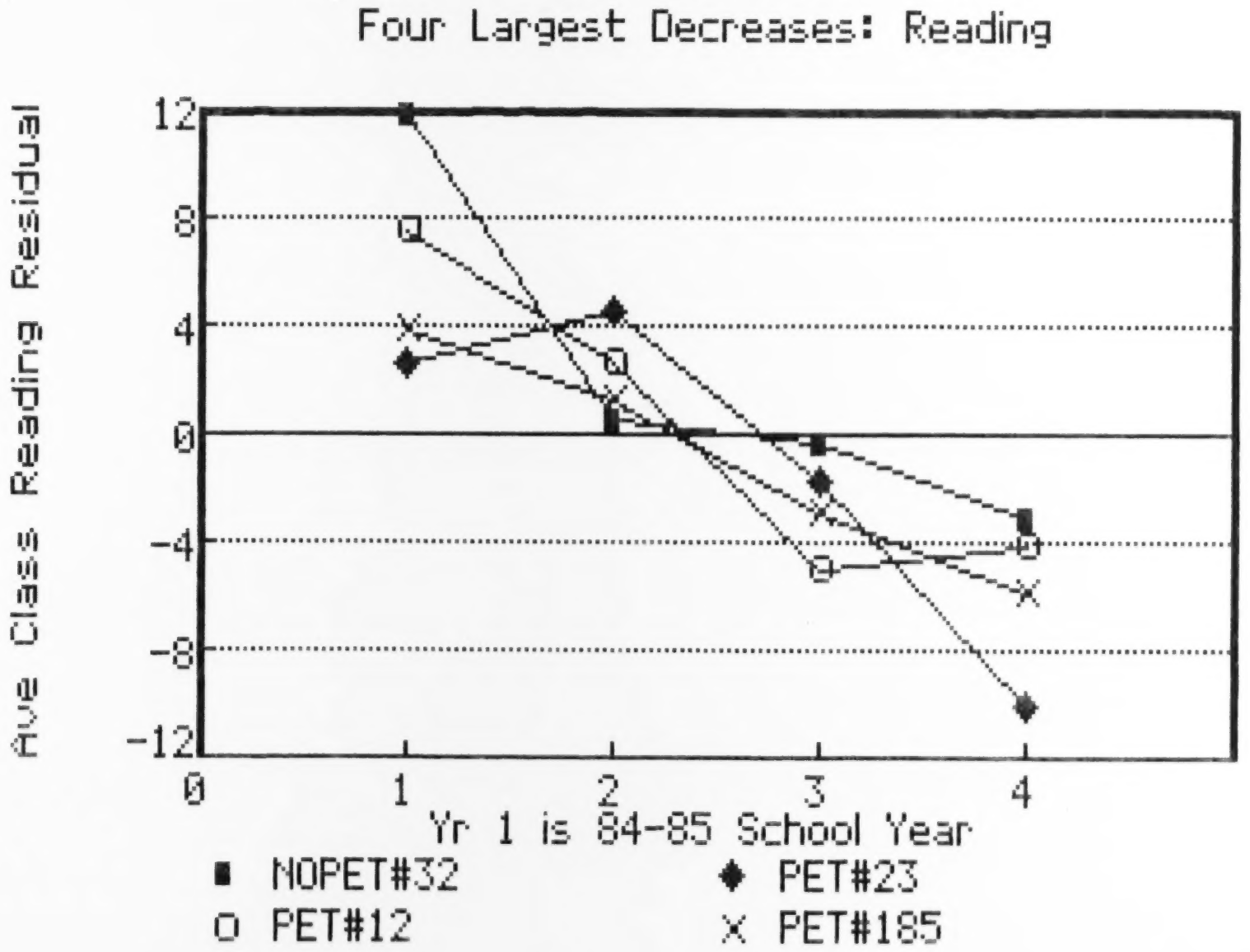
Four Largest Decreases: Reading
Ave Class Reading Residual
12
8
4
0
-4
-8
-12
0
1
2
3
4
Yr 1 is 84-85 School Year
NOPET#32
PET#23
PET#12
PET#185

Figure 3

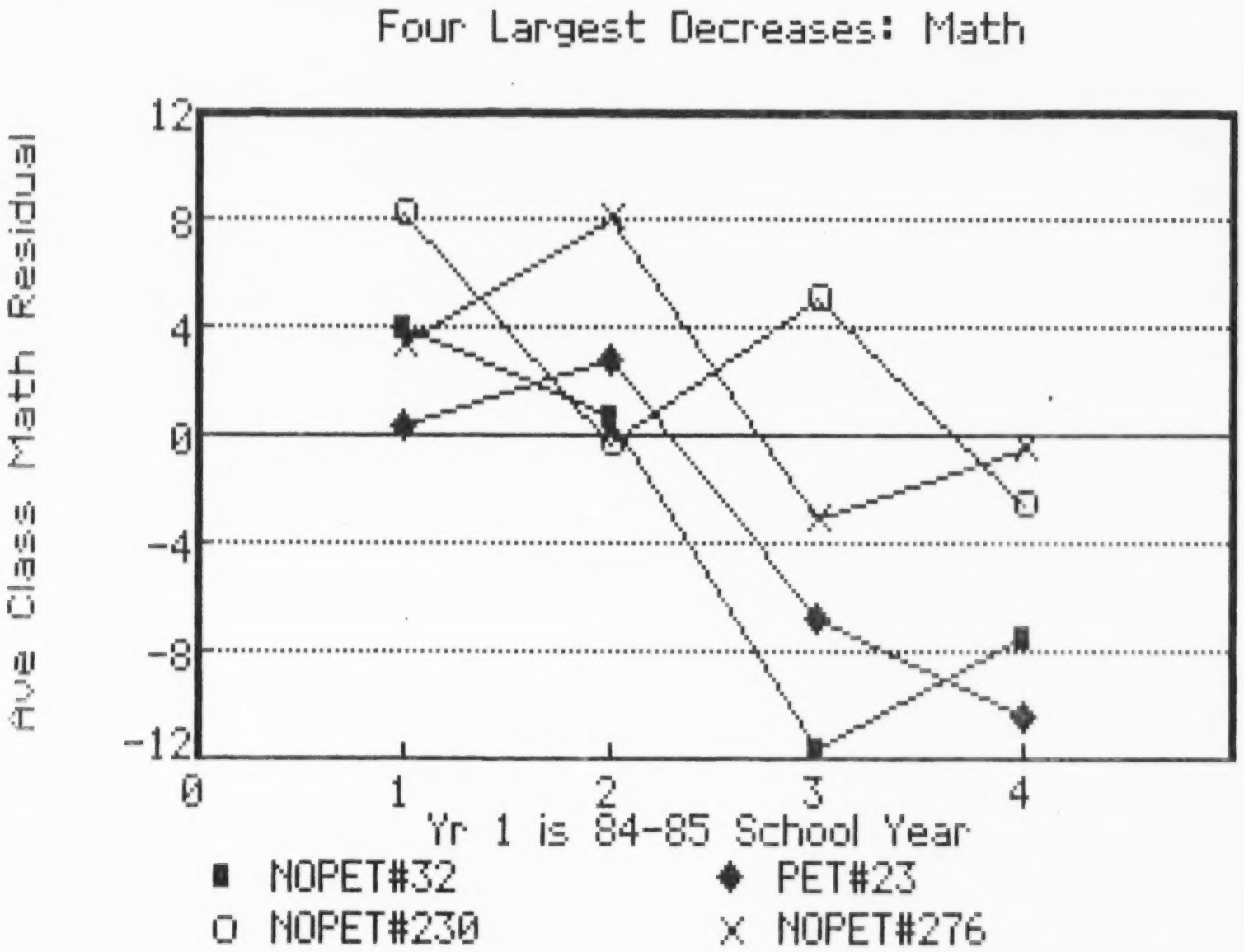
Four Largest Decreases: Math
Ave Class Math Residual
12
8
4
0
-4
-8
-12
0
1
2
3
4
Yr 1 is 84-85 School Year
NOPET#32
PET#23
NOPET#230
NOPET#276

Figure 4